

Statistique Générale

Sébastien Lê

([http://sebastien.ledien.free.fr/
Miscellaneous.html](http://sebastien.ledien.free.fr/Miscellaneous.html))

Les « Stat » à l'agro c'est...

...des projets

- *Hédonisme et catégorisation, application sur de données de vins*
- *Approches holistiques appliquées à des stimuli sonores: une étude sur des hauts parleurs de téléphones portables*

...des projets

- *Intimate Projective Mapping ou comment combiner contexte et émotions pour comprendre le consommateur, application à des rouges à lèvres*
- *La catégorisation libre, une nouvelle méthode de recueil de données*

...des projets

- *Modélisation de la température de la surface de la mer*
- *La malnutrition dans la région de Dakar*

...des prix



Trophées Syntec des Etudes

- Vous êtes étudiant de master en marketing et études de marché, **Participez à la nouvelle édition des Trophées des Etudes Marketing & Opinion.**
- Quatre catégories de prix sont proposées:
 - Mémoire de master recherche M2: **1500 euros**
 - Mémoire de master professionnel M2: **1500 euros**
 - Mémoire de master M1 (collectif ou individuel):**1500 euros**
 - Mémoire collectif: **2500 euros**

Résultats de la 1ère édition des Trophées Syntec des Etudes

- **Mémoire Master Collectif :**
- **L'influence de la marque sur la perception du consommateur?**
- Gaël Gauthier, Guilaine Perron, Cécile Perreau, Gaël Chevet

Résultats de la 2ème édition des Trophées Syntec des Etudes

- **Mémoire Master Collectif :**
- **Les consommateurs ont la parole ! Application à douze parfums de luxe**
- Mélanie Cousin, Maelle Penven, Mathilde Philippe, Marie Toularhouat

Résultats de la 3ème édition des Trophées Syntec des Etudes

- **Mémoire Master Collectif :**
- Quelle est l'influence de l'apport d'information telle que la liste des ingrédients et l'appellation du produit sur la perception du consommateur?
- Lucie DECRETON, Carole JEGOU, Laure LE PAIH, Magali MONNIER

Résultats de la 4ème édition des Trophées Syntec des Etudes

- **Mémoire Master Collectif :**
- "Une nouvelle méthode de tri : le tri hiérarchique"
- SCHMIED Camille, de la PORTBARRÉ Charles, MELHINGER Stéphanie, KEMARREC Cathy, HERSCHTAL Mélusine

Résultats de la 5ème édition des Trophées Syntec des Etudes

- Catégorie Master 1. **Affichage de l'éco-conception : approche quantitative d'un problème complexe.** Perrine Baudry et Coline Rondeau .
- Catégorie master 2, mémoire collectif. **Intimate projective mapping : comment le contexte et les états émotionnels peuvent aider à comprendre le comportement du consommateur.** Prix spécial du jury décerné à Maud Leportier, Lisa Defeyter et Charlène Benniza.

...des stages



- Recherche d'une méthode de caractérisation sensorielle du haricot vert surgelé de la marque Bonduelle.



- Participation au développement de nouveaux produits ou à l'optimisation de produits existants dans la gamme des fourrages/pâte à glacer/fourrage aqueux.



- Analyse et comparaison des résultats de tests consommateurs afin de recommander des méthodes permettant des résultats « stables ».

Dior

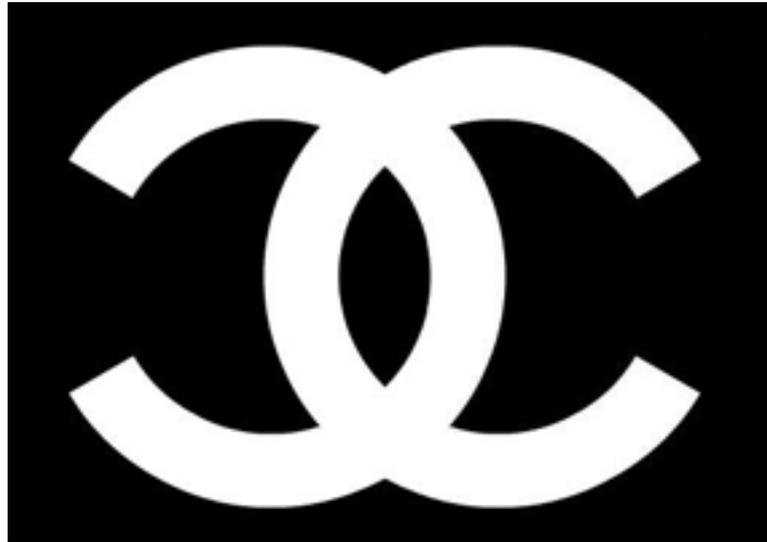
- Accompagner la mise en place d'un "Preference mapping" sur l'hydratation visant à mieux comprendre la perception de l'hydratation, sujet de fond pour les équipes de développement R&D.

L'ORÉAL

- Analyse de la fiabilité des données recueillies et réalisation de mapping de produits sur la base de données d'évaluation.

International Flavors & Fragrances

- Optimisation d'une approche sensorielle dans la mesure de la performance des parfums et des matières premières de parfumerie par un panel interne.



- Recherche d'une cartographie de produits (soins et maquillage).

Inserm



**Institut national
de la santé et de la recherche médicale**

- Identification statistique de médicaments associés aux risques d'accident de la circulation.



- Analyse des déterminants de l'exposition combinée de la pollution à des mélanges de pesticides.



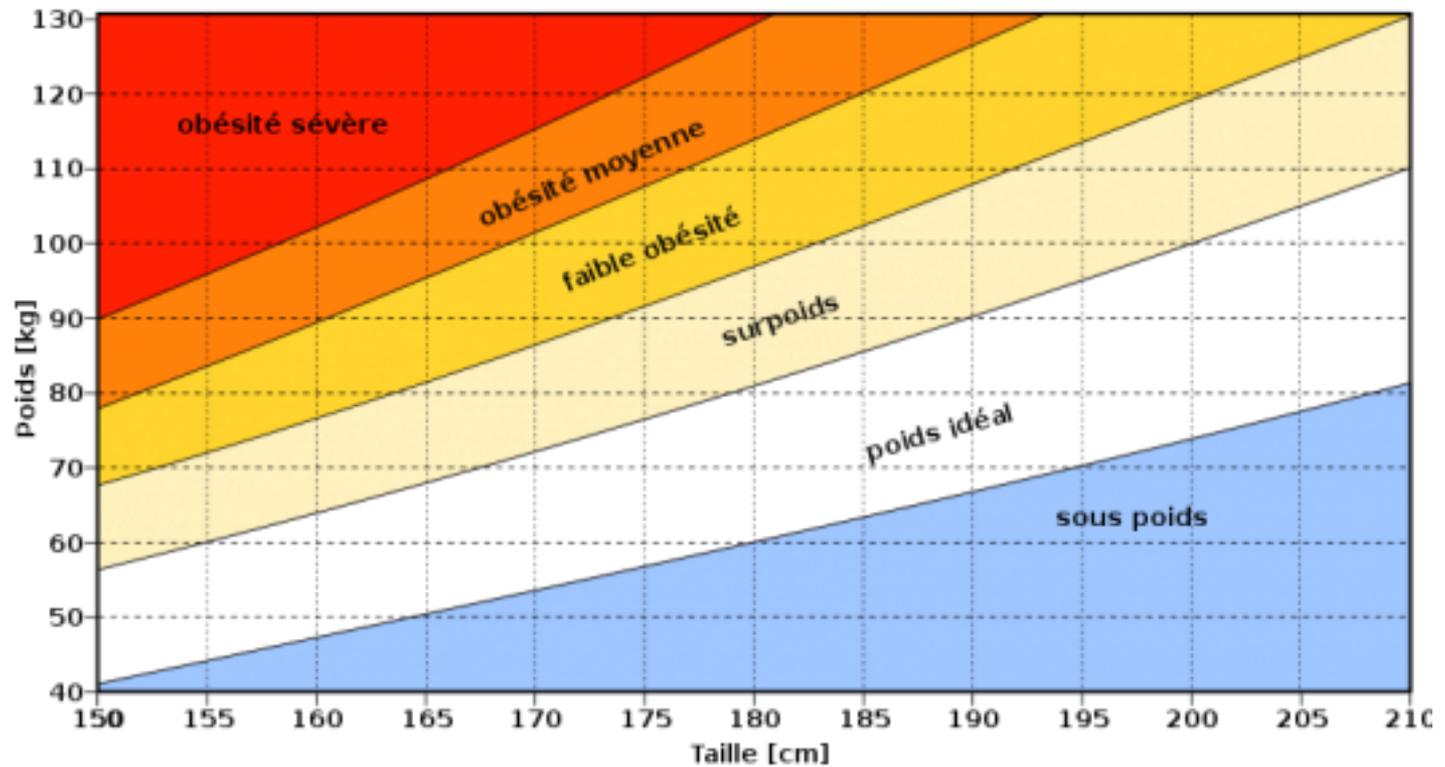
- Analyse statistique de données d'écologie microbienne sous R.
- Etude statistique de la réponse de la végétation en communauté d'alpage à différents patrons spatiaux de défoliation expérimentation.

PETIT PROBLÈME DE STATISTIQUE...

<http://bigbrowser.blog.lemonde.fr/2012/06/25/super-size-me-plus-dune-britannique-sur-deux-est-en-surpoids/>

- *"Un problème inquiétant."* C'est ainsi que le Dr Davina Deniszczyc définit à la BBC le résultat de l'étude qu'elle a menée avec d'autres chercheurs pour l'ONG caritative Nuffield Health.
- D'après les données portant sur 30 000 femmes, elle conclut que 52,5 % des femmes britanniques sont en surpoids, c'est-à-dire avec un indice de masse corporelle (IMC, obtenu en divisant le poids en kilogrammes par la taille en mètres au carré) supérieur au "poids idéal" (IMC entre 18,5 et 25) ; 16,2 % d'entre elles sont considérées obèses (IMC supérieur à 30).

<http://bigbrowser.blog.lemonde.fr/2012/06/25/super-size-me-plus-dune-britannique-sur-deux-est-en-surpoids/>



Tour de taille moyen de 84,9 cm, 85,5 cm pour les Françaises

- L'étude constate également que 57 % des femmes britanniques ont un tour de taille plus important qu'il ne devrait pour garantir une bonne santé. Le tour de taille moyen est de 84,9 cm, quand il devrait normalement être d'environ 80 cm. C'est dans le nord de l'Angleterre qu'il est le plus important, avec 87 cm en moyenne, contre 81,9 cm à Londres.
- Davina Deniszczyc met en garde contre les conséquences de la situation : "Cela peut ressembler à un sujet cosmétique, il n'est pas question des femmes rentrant dans les jeans serrés. La graisse concentrée autour de la taille peut conduire à d'importants problèmes de santé, comme des cancers du sein et l'infertilité."
- Les femmes britanniques ne sont pas les seules concernées : en France, selon [l'enquête ObEpi de Roche avec l'institut public Inserm, publiée en 2009](#), 26 % des Françaises (et 38,5 % des Français) sont considérées en surpoids, 15,1 % étant obèses (13,9 % chez les hommes). Concernant le tour de taille moyen, celui-ci est de 85,5 cm chez les Françaises, soit 5,7 cm de plus qu'en 1997 (79,8 cm). Celui des hommes est, quant à lui, passé de 91,3 cm à 94,8 cm.

Comment ont-ils fait?

- Sachant que la « population française » est estimée à 65 350 181 habitants (1^{er} Janvier 2012)
- Sachant que la « population française » est composée de 51.4% de femmes
- Sachant que chaque minute 1.010464 personne décède en France
- Sachant que chaque minute 1.577915 enfant naît en France

Démarche statistique

- Question
- Recueil
- Traitement
- Décision
- Retour éventuel à la question

Nature des variables mesurées

- Quantitatif
- Qualitatif

Nature du problème

- Statistique exploratoire
- Statistique inférentielle
- Population
- Echantillon

Nature du problème

- Statistique exploratoire
- Statistique inférentielle
- Population
- Echantillon
- **Démonstration avec R**

Suite à notre discussion...

- La question à laquelle je cherche à répondre m'amène à me situer soit dans un contexte **exploratoire**, soit dans un contexte **inférentiel**
- On ne peut pas prendre de décision sur la base d'une seule **moyenne**, il nous faut l'information sur la **variance** (ou celle sur l'écart type)

Suite à notre discussion...

- La distribution d'échantillonnage nous permet de comprendre les propriétés de nos estimateurs
- On distingue
 - estimation ponctuelle
 - estimation par intervalle

Suite à notre discussion...

La **moyenne** $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

est un estimateur « naturel » de

la **moyenne** μ (que je ne connaîtrai jamais)

\bar{X}_n a en plus de bonnes propriétés

Distribution d'échantillonnage

$$E(\bar{X}_n) = \mu$$

$$V(\bar{X}_n) = \frac{\sigma^2}{n}$$

Estimation par intervalle de confiance (1)

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0; 1)$$

$$\left[\bar{X}_n - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{X}_n + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

Rappels

La **variance** $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$

est un estimateur « naturel » de
la **variance** σ^2 (que je ne connaîtrai jamais)

On lui préférera $S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$

qui a de bonnes propriétés

Estimation par intervalle de confiance (2)

$$\frac{\bar{X}_n - \mu}{S'/\sqrt{n}} \sim T(n-1)$$

$$\left[\bar{X}_n - t_{1-\alpha/2}^{(n-1)} \frac{S'}{\sqrt{n}}; \bar{X}_n + t_{1-\alpha/2}^{(n-1)} \frac{S'}{\sqrt{n}} \right]$$

Estimation par intervalle de confiance (3)

$$V(\bar{X}_n) = \frac{\sigma^2}{n} \frac{N - n}{N - 1}$$

$$\left[\bar{X}_n - t_{1-\alpha/2}^{(n-1)} \frac{S'}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}; \bar{X}_n + t_{1-\alpha/2}^{(n-1)} \frac{S'}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}} \right]$$

Suite à notre discussion...

La **variance** $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$

est un estimateur « naturel » de
la **variance** σ^2 (que je ne connaîtrai jamais)

On lui préférera $S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$

qui a de bonnes propriétés

Estimation par intervalle de confiance

$$\frac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\left[\frac{(n-1)S'^2}{q_{1-\alpha/2}^{(n-1)}}; \frac{(n-1)S'^2}{q_{\alpha/2}^{(n-1)}} \right]$$

Fin de la 1^{ère} semaine

Cours 1

TD1 et TD2

Message important!

- Il existe des livres qui correspondent à ce qui est fait en cours et en TD, vous pouvez les acheter à tarif préférentiel au secrétariat de mathématiques appliquées (bât. 24)
- Vous êtes évalués par contrôle continu (autrement dit, il y aura un ou plusieurs TD évalués)
- Le jour de l'évaluation vous avez droit à une feuille de formules (format A4)
- Pensez à prendre vos calculatrices

Tour de taille moyen de 84,9 cm, 85,5 cm pour les Françaises

- L'étude constate également que 57 % des femmes britanniques ont un tour de taille plus important qu'il ne devrait pour garantir une bonne santé. Le tour de taille moyen est de 84,9 cm, quand il devrait normalement être d'environ 80 cm. C'est dans le nord de l'Angleterre qu'il est le plus important, avec 87 cm en moyenne, contre 81,9 cm à Londres.
- Davina Deniszczyc met en garde contre les conséquences de la situation : "Cela peut ressembler à un sujet cosmétique, il n'est pas question des femmes rentrant dans les jeans serrés. La graisse concentrée autour de la taille peut conduire à d'importants problèmes de santé, comme des cancers du sein et l'infertilité."
- Les femmes britanniques ne sont pas les seules concernées : en France, selon [l'enquête ObEpi de Roche avec l'institut public Inserm, publiée en 2009](#), 26 % des Françaises (et 38,5 % des Français) sont considérées en surpoids, 15,1 % étant obèses (13,9 % chez les hommes). Concernant le tour de taille moyen, celui-ci est de 85,5 cm chez les Françaises, soit 5,7 cm de plus qu'en 1997 (79,8 cm). Celui des hommes est, quant à lui, passé de 91,3 cm à 94,8 cm.

Test d'hypothèse

- Se comparer par rapport à une valeur cible
- Comparer deux moyennes entre elles
- Il y a deux hypothèses, mais en réalité, une seule des deux est testée

Statistique de test

$$T = \frac{\bar{X} - \mu_0}{\hat{\sigma}(\bar{X})}$$

$$T \stackrel{H_0}{\sim} T(n - 1)$$

Région de rejet

Adéquation à une valeur cible

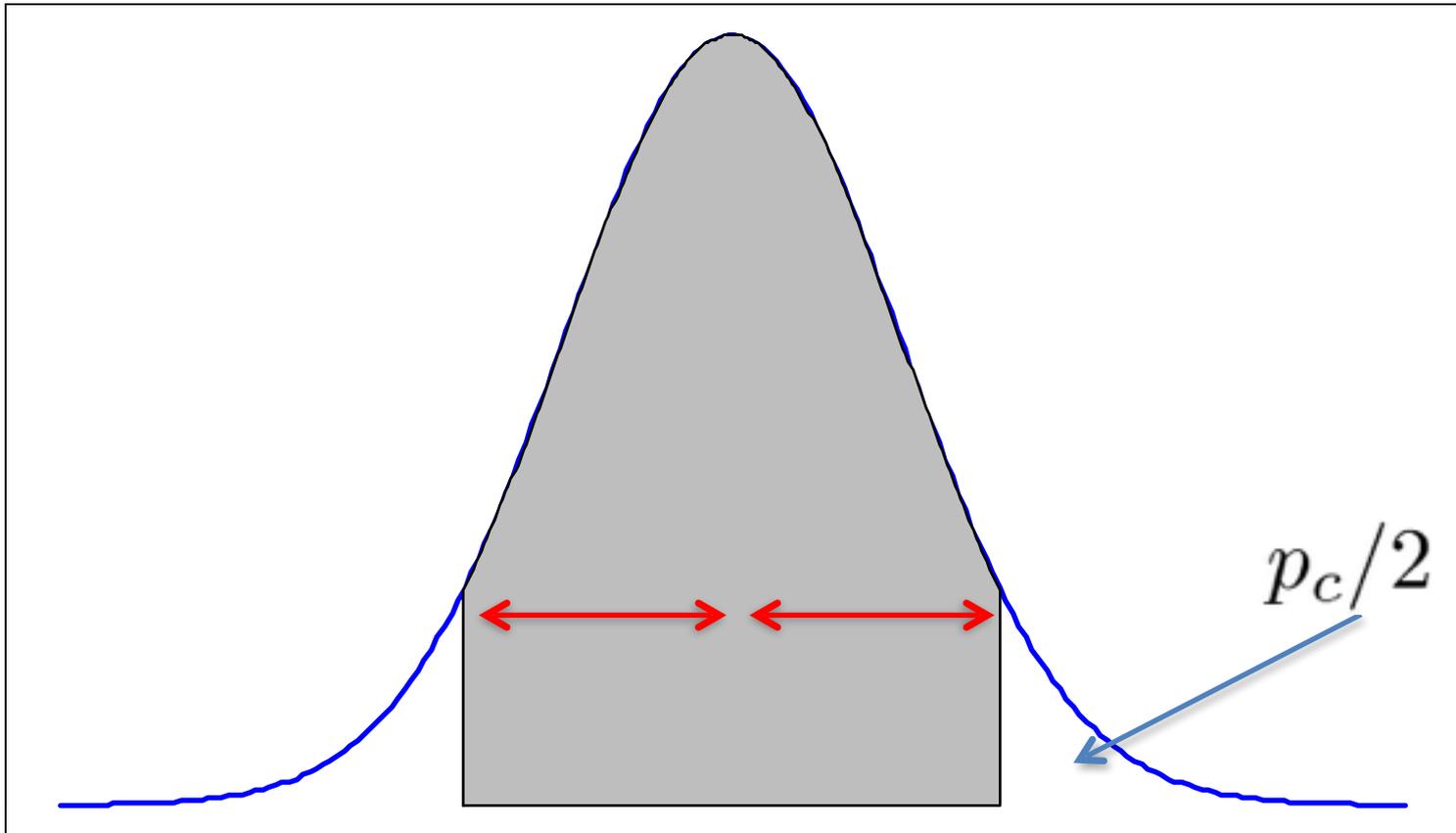
$$\bar{X} \in] -\infty, \mu_0 - t_{1-\alpha/2}^{(n-1)} \frac{S'}{\sqrt{n}}] \cup [\mu_0 + t_{1-\alpha/2}^{(n-1)} \frac{S'}{\sqrt{n}}, +\infty [$$

Probabilité critique

- C'est la probabilité d'obtenir, par tirage d'un échantillon au hasard sous l'hypothèse H_0 , un écart au moins aussi grand que celui effectivement observé

Probabilité critique

En rouge: écart observé



Comparaison de deux moyennes

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Comparaison de deux moyennes

On suppose ici les variances connues

- Loi d'échantillonnage de la différence de deux moyennes

$$D = \bar{X}_1 - \bar{X}_2$$

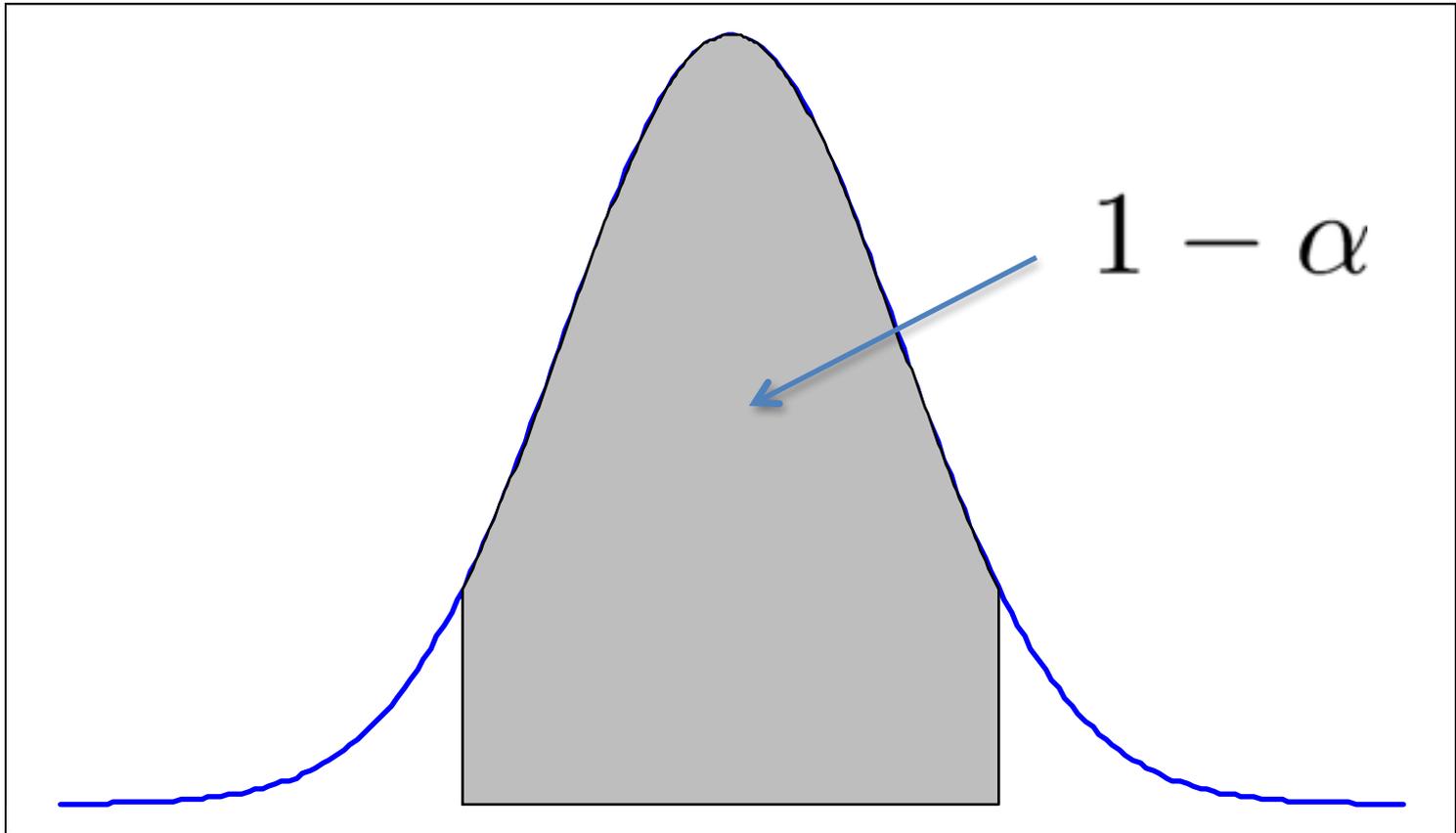
$$\sigma_D^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Dans le cas où l'hypothèse H_0 est vraie

$$D \sim N(0; \sigma_D^2)$$

$$D/\sigma_D \sim N(0; 1)$$

Région de rejet



$$u_{\frac{\alpha}{2}}$$

$$u_{1 - \frac{\alpha}{2}}$$

Probabilité critique

- C'est la probabilité d'obtenir, par tirage de deux échantillons au hasard sous l'hypothèse H_0 , un écart de moyennes au moins aussi grand que celui effectivement observé

Erreurs et risques

- Rejeter l'hypothèse nulle alors qu'elle est vraie (i.e. décider à tort de rejeter H_0), consiste à voir des différences là où il n'y en a pas. C'est l'erreur dite de première espèce
- La probabilité associée à cette erreur est appelée le « risque de 1^{ère} espèce »

α

Erreurs et risques

- Accepter l'hypothèse nulle alors que c'est l'alternative qui est vraie (i.e. décider à tort de ne pas rejeter H_0), consiste à ne pas voir une différence qui existe. C'est l'erreur dite de seconde espèce
- La probabilité associée à cette erreur est appelée le « risque de 2^{nde} espèce »

β

Erreurs et risques

		Décision	
		Accepter H_0	Rejeter H_0
H_0 vraie	$1 - \alpha$	α	
H_1 vraie	β	$1 - \beta$	

Puissance

- Fonction qui dépend de H_1
- D'autant plus puissant que la différence est grande
- Aptitude à mettre en évidence une différence entre deux moyennes (par exemple)

$$1 - \beta$$

Comparaison de deux moyennes

On suppose ici les variances inconnues, mais égales

$$\sigma_D^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\sigma_D^2 = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

Comment estimer la variance?

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_1'^2 + (n_2 - 1)s_2'^2}{n_1 + n_2 - 2}$$

Dans le cas où l'hypothèse H_0 est vraie,

$$\text{i.e. } \mu_1 = \mu_2$$

$$D/\hat{\sigma}_D \sim T(n_1 + n_2 - 2)$$

Comparaison de deux variances

$$\frac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\frac{(n_1-1)S_1'^2}{\sigma_1^2} \sim \chi^2(n_1-1)$$

$$\frac{(n_2-1)S_2'^2}{\sigma_2^2} \sim \chi^2(n_2-1)$$

Comparaison de deux variances,
loi si H_0 est vraie

$$\frac{S_1'^2}{S_2'^2} \sim F_{\substack{(n_1 - 1) \\ (n_2 - 1)}}$$